DIF Registry and Catalog Working Group

NOAA Coastal Services Center
LINKING PEOPLE, INFORMATION, AND TECHNOLOGY

Overview

Background

- Charge to Working Group
- Members
- Context

Relevant Technologies and Projects

- Opportunities and Limitations

Working Group Recommendations

Working Group Charge

Review the DIF requirements related to Registries and Catalogs, survey relevant technologies and projects, and produce recommendations for the IPT.

Working Group Members

John Ulmer

Matthew Howard

Eric Bridger

Lewis McCulloch

Daniel Martin

Carmel Ortiz

Alex Birger

Jeff de La Beaujardiere

Some Context: Requirements

High Level or Coarse Resolution Data/Service Discovery

- Must be available and responsive to update by registrar.
- Inclusion indicates data provider meets qualification criteria.
- Provides human interface to SOA – high level descriptions of available services at the Registry level.
- Provides machine interface to SOA – readable and actionable high level descriptions of available services at the Registry level.
- Describes service interface.
- Describes general temporal and spatial extent of holdings and available *observedProperties.*

Some Context: Requirements (cont'd.)

Low Level or Fine Resolution Data Discovery/Examination

- Supports examination of candidate data sets for appropriate-use considerations (appropriate QA/QC, adequate resolution, appropriate/consistent measurement techniques, calibration, etc.)
- Content must originate at the data provider.
- Both data sets and data services must be documented with FGDC or ISO metadata.

Some Context: Requirements (cont'd.)

Catalog

Repository of controlling documents such as schemata, data dictionaries, process control documentation, etc.

More Context: Catalog or Registry?

What are we talking about?

From eb/RIM profile of OGC CSW Spec.

> *The terms 'catalog' and 'registry' are often used interchangeably, but, the following distinction is made in this application profile: a registry is a specialized catalogue that exemplifies a formal registration process... A registry is typically maintained by an authorized registration authority who assumes responsibility for complying with a set of policies and procedures for accessing and managing registry content.*

More Context: Drivers

Short-term Drivers

- Limited resources (money, people, time).
- Modest functional needs.
- Small community of involvement (single federal agency + regions)
- Consideration of transition to long-term strategies

More Context: Drivers

Long-term Drivers

- More resources possible, but still likely limited
- Expanding functional needs.
  - Higher resolution data (temporal and spatial).
  - Additional observed properties.
  - Rigorous system monitoring and reaction.
  - Continuity of Operations requirements at the central and data provider levels.
- Span numerous federal and non-federal data providers

Relevant Technologies & Projects

GEOSS

- Has a registry of services.
- NDBC and CO-OPS services are registered.
- Lacks some fields identified by the WG as desired.

Geodata.gov

- Good repository for data set level and service level metadata records.
- Interaction with GOS is not trivial.

NOS Data Explorer – An NOS Data Portal

- The DE project is beginning an effort to rewrite their portal software based on the ESRI Portal Tool Kit.
- Their metadata domain is limited to NOS.
- Their software maybe available and appropriate for use by the DIF once completed.

IOOS Observation Registry

- Enjoys wide regional support.
- Data model overlaps the DIF Registry needs significantly.
- Future is unknown.

Relevant Technologies & Projects

OGC Catalog Service for the Web (CSW)

- Certainly relevant, but, is a high level fairly abstract specification that will require significant development of implementation level detail.
- eb/RIM adds some specificity to CSW, but, is not trivial and will require significant investment to determine potential utility.

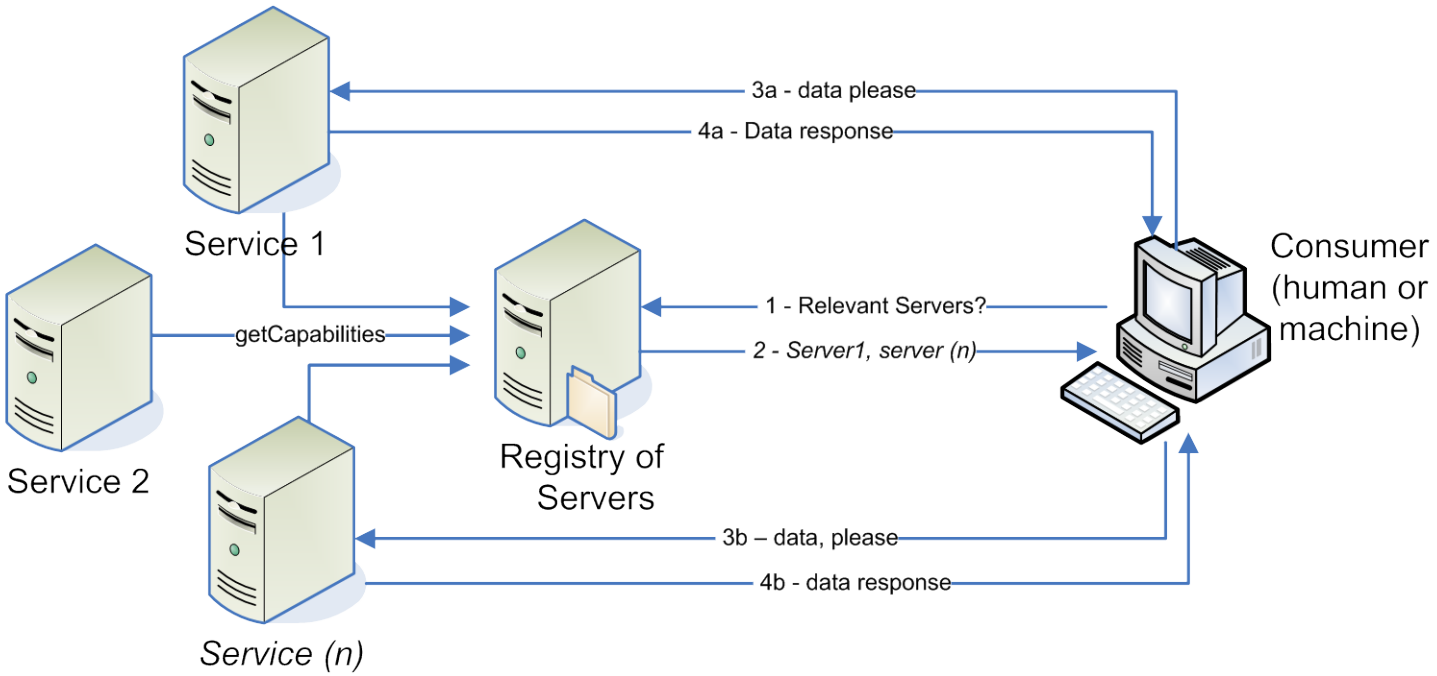
Relevant ISO Metadata Specifications

- ISO 19135 – defer to Metadata Working Group
- ISO 19115 – defer to Metadata Working Group
- ISO 19119 - Directly relevant to the DIF Service Oriented Architecture
  - Offers data model and vocabulary for DIF implementation
  - Requires some content beyond what RoS working group identified for the Registry.

NOAA's GEO-IDE FY 2010 Project (workshop to follow IOOS IPT)

- Project is working on a unified access framework for NOAA's gridded datasets utilizing THREDDS
- Project just getting started so opportunity is there to coordinate and leverage development efforts

DIF SOA



Services may be of: SOS, DAP, THREDDS, or WCS.

Registry Content

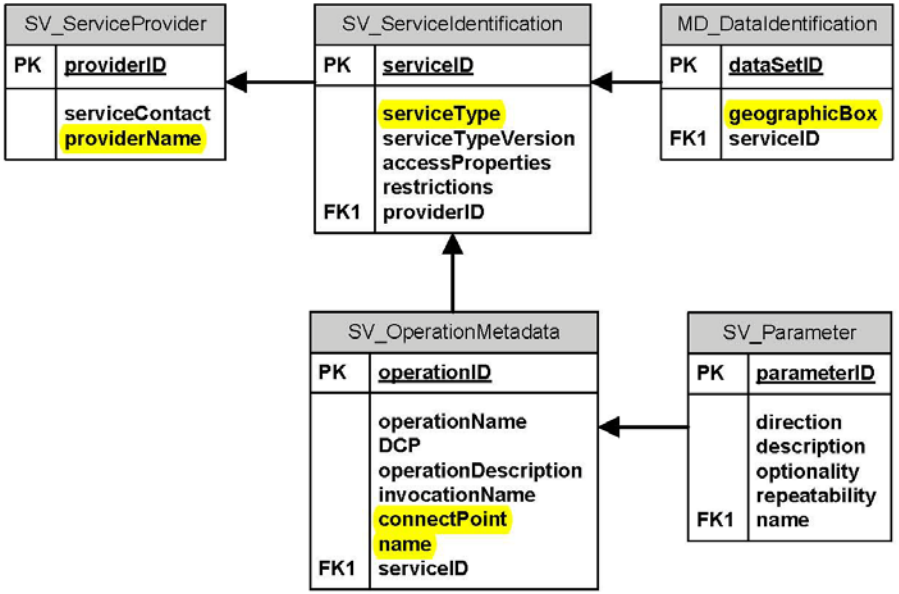
Fields Adopted from the Observation Registry Record Schema

- operator
- operatorURI
- observedProperty(ies)
  - startDate
  - endDate
  - (location)
- serviceDescription
  - serviceType (DAP, SOS, WCS, WMS, other?)
  - serviceURI

Location elements in Obs. Registry are sensor specific. RoS will need bbox of all sensors available via the service.

Registry Content in ISO 19119

Obs Registry Element	ISO 19119 Element
operator	SV_ServiceProvider : providerName
operatorURI	n/a
observedProperties	SV_parameter : Name
startDate	n/a
endDate	n/a
serviceDescription	SV_ServiceIdentification : serviceType
serviceURI	SV_OperationMetadta : connectionPoint
Bounding box	MD_DataIdentification: geographicBox



NOAA Coastal Services Center
LINKING PEOPLE, INFORMATION, AND TECHNOLOGY

Conceptual Registry Record Structure

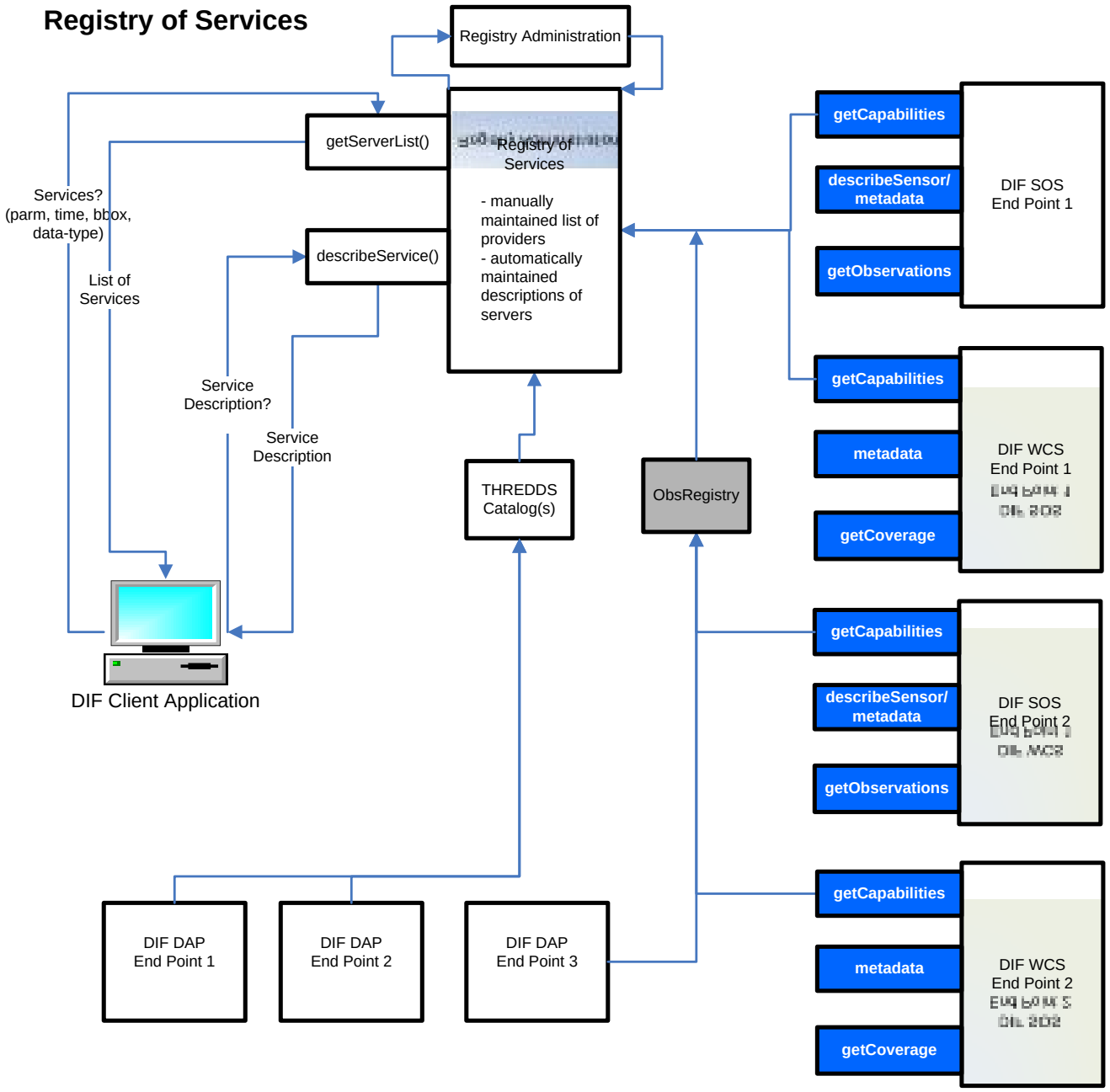
```
<?xml version="1.0" encoding="UTF-8"?>
<DIFDataProviders
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:noNamespaceSchemaLocation="RoS_Schema_Example_20090805.xsd">
 <registrant>
  <providerName>NDBC</providerName>
  <observedProperties>
   <observedProperty>
    <name>Winds</name>
    <startDate> 2003-07-04T00:00:00Z</startDate>
    <endDate/>
    <geographicBox/>
   </observedProperty>
   <observedProperty>
    <name>Salinity</name>
    <startDate> 2003-07-10T00:00:00Z</startDate>
    <endDate/>
    < geographicBox />
   </observedProperty>
  </observedProperties>
  < serviceIdentification >
   <serviceType>SOS</serviceType>
   <connectPoint>http://url.to.service.endpoint</connectPoint>
  </ serviceIdentification >
 </registrant>

 …
```

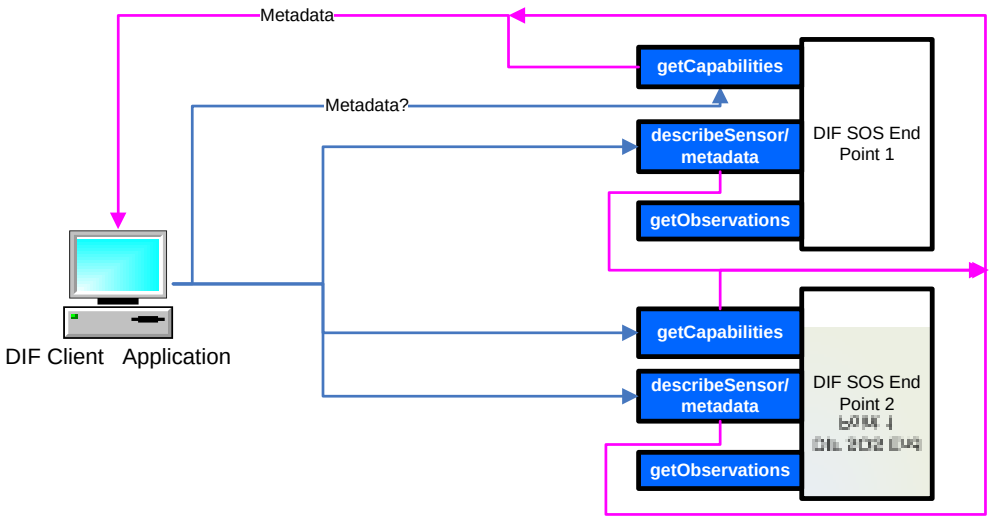
NOAA Coastal Services Center
LINKING PEOPLE, INFORMATION, AND TECHNOLOGY

Conceptual Registry Record Structure

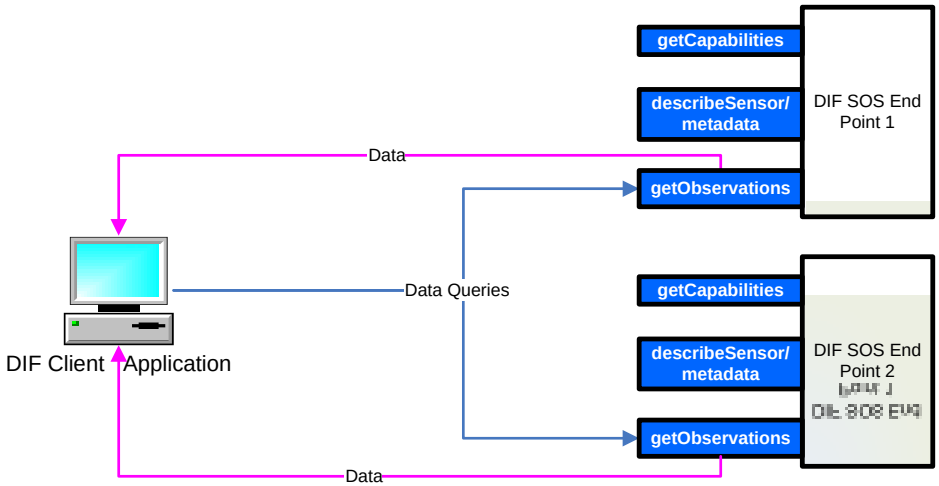
```
…
 <registrant>
  <providerName>CO-OPS</providerName>
  <observedProperties>
   <observedProperty>
    < name >Winds</ name >
    <startDate> 2003-07-04T00:00:00Z</startDate>
    <endDate/>
    < geographicBox />
   </observedProperty>
  </observedProperties>
  <serviceIdentification>
   <serviceType>DAP</serviceType>
   <connectPoint>http://url.to.service.endpoint</connectPoint>
  </ serviceIdentification >
 </registrant>
 <registrant>
  <providerName>AOOS</providerName>
  <observedProperties>
   <observedProperty>
    < name >Winds</ name >
    <startDate> 2003-07-04T00:00:00Z</startDate>
    <endDate/>
    < geographicBox />
   </observedProperty>
  </observedProperties>
  <serviceIdentification>
   <serviceType>WCS</serviceType>
   <connectPoint>http://url.to.service.endpoint</connectPoint>
  </serviceIdentification>
 </registrant>
</DIFDataProviders>
```


Registry of Services
Registry Administration
getServerList()
describeService()
Registry of Services
- manually maintained list of providers
- automatically maintained descriptions of servers
Services? (parm, time, bbox, data-type)
List of Services
Service Description?
Service Description
DIF Client Application
THREDDS Catalog(s)
ObsRegistry
getCapabilities
describeSensor/ metadata
getObservations
DIF SOS End Point 1
getCapabilities
metadata
getCoverage
DIF WCS End Point 1
getCapabilities
describeSensor/ metadata
getObservations
DIF SOS End Point 2
getCapabilities
metadata
getCoverage
DIF WCS End Point 2
DIF DAP End Point 1
DIF DAP End Point 2
DIF DAP End Point 3

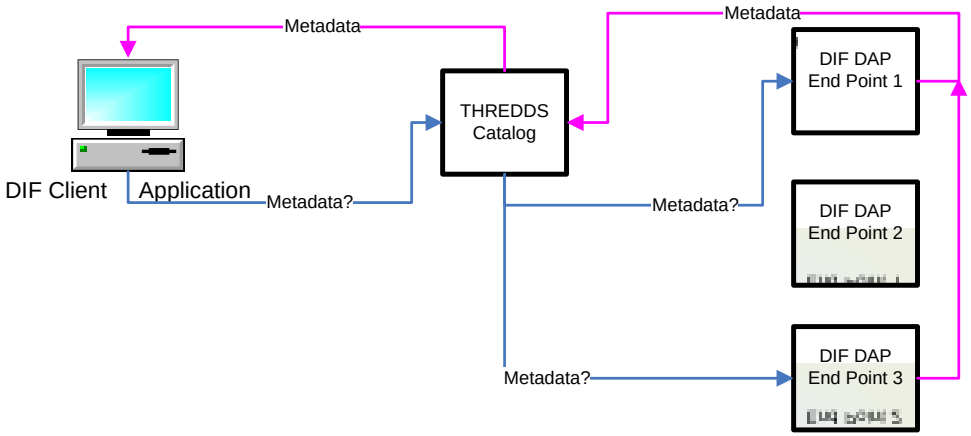
Client Query for Metadata (SOS)



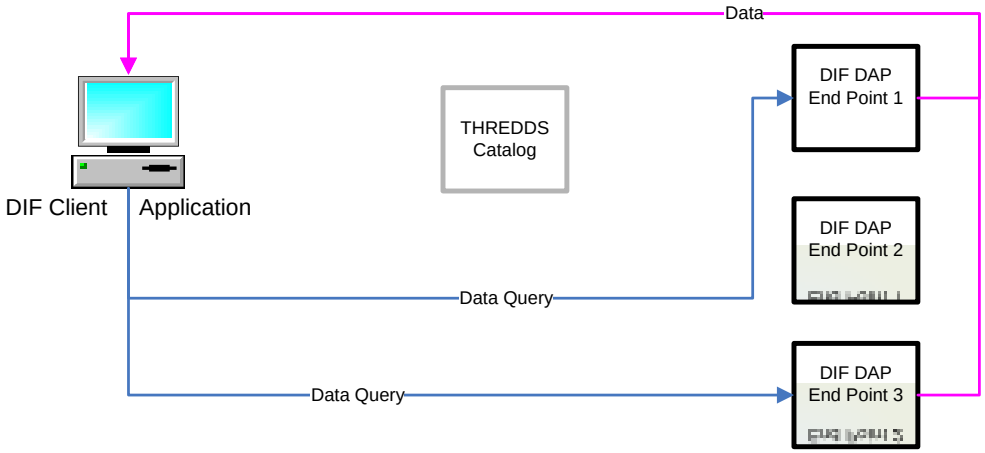
Data Request (SOS)



Client Query for Metadata (DAP)



Data Request (DAP)



Working Group Recommendations

Establish an administrative process and registrar to manage DIF data providers. Elements would include:

- A standardized testing process.
- Articulated requirements for compliance.
  - Minimum content for each service type.
  - Minimum level of continuity of service.
  - Minimum server performance.
- Communication mechanisms with existing and candidate data providers.
- Monitoring of data provider performance for compliance.
- Standardization of interfaces to underlying catalogs (THREDDS and Obs Registry).

Working Group Recommendations

Catalog

Consolidate administrative, process control, schema, data dictionaries, and other documents relevant to the design and operation of the DIF SOA into a single location (physical or virtual).

Working Group Recommendations

Implement a Simple Registry of Services

- Online XML Listing of data providers.
  - Manually maintained list of data providers as determined by the registrar.

- Registry of Services
  - Automatically populated listing of services available from those providers pulling data from Obs. Registry and THREDDS catalogs.
  - Supports queries on ***observedProperty, time range, bounding box, and dataType*** (in-situ, model, remotely sensed).
  - Returns a list of servers containing relevant data.
  - Returns a description of a specific service.

Working Group Recommendations

- Transfer this work to a permanent body for further development.
- Map and bridge dependencies between DIF Service Registry and THREDDS and Observation Registry content.
- Continue collaboration with GEOSS.
- Continue collaboration with NOAA's GEO-IDE initiative
- Continue collaboration with the Observation Registry.
- Support a proof-of-concept level project to fully explore the the OGC Catalog Service for the Web (CSW) and eb/RIM.
- Incorporate relevant ISO metadata standards into DIF/DMAC development.